Available online at www.sciencedirect.com
SCIENCE DIRECT®
Technological Forecasting & Social Change
70 (2003) 735–758
NORTH-HOLLAND
Technological Forecasting and Social Change

R&D cluster quality measures and technology maturity

Robert J. Watts[a,*], Alan L. Porter[b,c]

[a]*U.S. Army Tank-Automotive and Armaments Command, Advanced Vehicle Technologies, Warren, MI 48397-5000, USA*
[b]*Search Technology, Inc., Suite 230, 4960 Peachtree Industrial Boulevard, Norcross, GA 30071-1580, USA*
[c]*Technology Policy and Assessment Center, Georgia Tech., Atlanta, GA 30332-0205, USA*

Received 24 June 2002; accepted 7 August 2002

Abstract

"Innovation indicators" strive to track the maturation of an emerging technology to help forecast its prospective development. One rich source of information is the changing content of discourse of R&D, as the technology progresses. We analyze the content of research paper abstracts obtained by searching large databases on a given topic. We then map the evolution of that topic's emphasis areas.

The present research seeks to validate a process that creates factors (clusters) based on term usage in technical papers. Three composite quality measures—cohesion, entropy, and *F* measure—are computed. Using these measures, we create standard factor groupings that optimize the composite term sets and facilitate comparisons of the R&D emphasis areas (i.e., clusters) over time.

The conceptual foundation for this approach lies in the presumption that domain knowledge expands and becomes more application specific in nature as a technology matures. We hypothesize implications for this knowledge expansion in terms of the three factor measures, then observe these empirically for the case of a particular technology—autonomous navigation. These metrics can provide indicators of technological maturation.
Published by Elsevier Science Inc.

Keywords: R&D cluster quality measures; Technology maturity; Innovation indicators

* Corresponding author.
E-mail address: wattsb@tacom.army.mil (R.J. Watts).

0040-1625/02/$ – see front matter. Published by Elsevier Science Inc.
doi:10.1016/S0040-1625(02)00355-4

1. Background: TECH OASIS and R&D profiling

Technology managers have many reasons to want to gauge how rapidly a technology of interest is progressing toward applications. Most organizations, private and public, need to assess external technological developments to determine how they can gain from these (e.g., via joint development activities, licensing). The good news—there is tremendous information available relating to scientific and technological development activities. In particular, large, publicly accessible databases compile such information and make it electronically accessible (for a price). Two such databases do a fine job of abstracting a major portion of the world's open engineering R&D literature—INSPEC (Institute of Electrical Engineers, UK: http://www.iee.org/Publish/INSPEC/) and *EI Compendex* =http://www.ei.org/eicorp/eicorp?menu = engineeringvillage2menu&display = engineeringvillage2). Together, they add about 500,000 abstracts of conference papers and journal articles annually.

The bad news? The quantity of information available on a given technology exceeds our traditional mechanism of digesting this—namely, reading. For instance, were you to want to keep track of developments in fuel cells, you would confront about 50,000 abstracts in the leading five or so databases. What to do? Given this need for information about emerging technologies and the abundance of such information in electronic form, we need to devise tools to exploit this information to help assess current developmental status and future prospects of a given technology.

Work on text mining is extremely active. This draws on efforts under several labels, including "KDD" (Knowledge Discovery in Databases—cf. http://www.cs.cmu.edu/~dunja/WshKDD2000.html, http://www.cs.biu.ac.il/~feldman/ijcai-workshop%20cfp.html), and bibliometrics (counting of bibliographic activity—cf. sistm.web.unsw.edu.au/conference/issi2001).

The Technology Opportunities Analysis of Scientific Information System (*TECH OASIS*) is a software tool that enables "text mining" of fixed field literature abstract files. That is, it counts the occurrences of particular terms, making it easy to list the most frequent authors, organizations researching the topic, terms used in the abstracts, etc. Such lists can be crossed with each other to create matrices. For instance, one might cross the leading keywords against the date of publication to see which keywords are most prevalent in recent years. van Raan [1] has called these "one-dimensional" (lists) and "two-dimensional" (matrix) analyses. One can go further to study interrelationships based on co-occurrence of terms. For example, it might be of interest to note which authors publish with each other. Alternatively, one could group terms, such as keywords, to see which tend to appear in the same abstracts.

TECH OASIS supports the performance of technology assessments by automating the profiling of open-source R&D. *TECH OASIS* has been used to serve the process of "innovation forecasting," by applying bibliometric analyses to augment and enhance traditional technology forecasting techniques [2].

TECH OASIS has been developed under the joint sponsorship of the Defense Advanced Research Projects Agency (DARPA) and the U.S. Army Tank-Automotive and Armaments Command (TACOM). The technology opportunities analysis (TOA) concept originated at

Georgia Tech's Technology Policy and Assessment Center (TPAC). TPAC strives to facilitate analyses of technological innovations [3,4] (see also http://tpac.gatech.edu). *TECH OASIS*, named *VantagePoint* for the commercial market, has been developed as a Windows-based software suite of tools that combines bibliometrics with content analysis [5]. *TECH OASIS* development represents a collaborative program, involving Search Technology, as the prime contractor, and subcontractors, Georgia Tech TPAC and Intelligent Information Services Corporation (IISC).

The TOA process entails these main steps:

1. Search and retrieve text information, typically from large abstract databases on a particular subject. In this paper, we analyze abstracts retrieved to capture research related to "autonomous navigation."
2. Clean the data and generate basic analyses.
3. Profile the resulting research domain [6]. *TECH OASIS* applies a combination of machine learning, statistical analyses enhanced by computational linguistics, fuzzy analysis, and principal components analysis (PCA), among others, to analyze literature abstracts. Profiling may focus on documents (e.g., "bucketing" documents into related, manageable groups [5,7]). Alternatively, it may focus on concepts (e.g., PCA to group related terms as conceptual clusters [8,9]). A third choice is a combination—seeking to link documents to concepts (e.g., relevance scoring [10]). Conceptual distinctions and methods are discussed further elsewhere [11].
4. Extract latent relationships. TECH OASIS applies iterative PCA to uncover links among terms and underlying concepts (cf. examples on http://tpac.gatech.edu [5,7–9,12]).
5. Represent relationships graphically. Generation of "maps," as applied in this research, is elaborated elsewhere [13].
6. Interpret the prospects for successful technological development. This typically entails integrating the bibliographic search set analyses with expert domain knowledge (interviews) [8].

The TOA process strives to create knowledge from a "body" of literature beyond that obtainable by digesting individual pieces. Retrieved text is treated as data [14]. Text is parsed into informative units, counted, and patterns uncovered that can speak to information analysts' interests and management needs.

2. The present research case: autonomous navigation

This paper focuses on extracting latent relationships through PCA and representing the derived relationships graphically. PCA, an inductive approach, does not impose groupings, but instead elicits them from the data. The PCA factor map analysis, a partly automated process, elicits relationships based on "co-occurrence" information. Co-occurrence reflects the pattern of terms occurring together. If two terms occur together in the records more frequently than expected, there is a presumption of relationship between them. The terms analyzed in this study are from the descriptors, or keywords, field of *INSPEC* and *EI Compendex* abstracts. The descriptors field for each abstract generally contains about five to eight terms that were generated to reflect the contents of the abstracted research paper. PCA of the descriptors should, therefore, generate factors (groupings of terms) that depict domain knowledge of the set of research papers under study.

What is being analyzed? This research documents the analysis of 1629 *INSPEC* and 1091 *EI Compendex* abstracts of technical papers published on "autonomous navigation" between 1987 and 2001. The search strategy used to retrieve the literature abstracts was "autonomous (adjacent to) (navigation or vehicle(s))." Roughly equal "number-of-record" periods were separated—i.e., 1987–1990 (279 records), 1991–1993 (248 abstracts), 1994–1995 (306 records), 1996–1997 (288 abstracts), 1998–1999 (284 abstracts), and 2000–2002 (223 records)—for the *INSPEC* "autonomous navigation" documented research. The 1091 *EI Compendex* autonomous navigation abstracts were subdivided into the same periods, with the following periods' record breakout: 171 abstracts in 1987–1990, 164 in 1991–1993, 197 in 1994–1995, 197 in 1996–1997, 182 in 1998–1999, and 180 in 2000–2001.

The next section explains the quality measures used to determine the standard factor or cluster groupings. We then discuss the use of these measures to assess technology maturity. Later, we present the findings for the autonomous navigation research, followed by a discussion of future research directions for the proposed innovation indicators.

3. Three criteria for term factors

The leading keywords (descriptors) compiled for the autonomous navigation abstract records represent the content of the full documents abstracted. *TECH OASIS* has a process that applies a semiautomated version of PCA, a basic form of factor analysis. Henceforth, we refer to the resulting clusters of terms that are so grouped as "factors." The factors, derived from the analyzed descriptors, should reflect domain knowledge as it builds over the time periods.

The resulting factors are automatically tabulated and depicted in a standard factor map display representation. Figs. 1 and 2 depict the factor maps derived from the leading keywords in "autonomous navigation" research papers. Fig. 1 derives from the *INSPEC* records for 1987–1990; Fig. 2, for *INSPEC* records for 1991–1993. Consider Figs. 1—14 factors are shown.[1] Each represents a group of keywords that tend to occur together in the abstract records. The "aerospace control" factor (upper, center) consists of two high-loading keywords, "aerospace control" and "space vehicles." The software has an algorithm to distinguish those more highly correlated keywords from the others (PCA actually calculates the relationship between every keyword in the analysis and each factor constructed there-

[1] The factor map algorithm uses an absolute value, descriptor loading factor, threshold to define the existence (i.e., relevance) of a derived cluster group. The descriptor loading factors for one factor in the 13-factor analysis, depicted in Fig. 1, exceeded this threshold in both the positive and negative ranges of the loading-factors. Therefore, 14 factors were generated for the 279 autonomous navigation abstracts from 1987 to 1990.

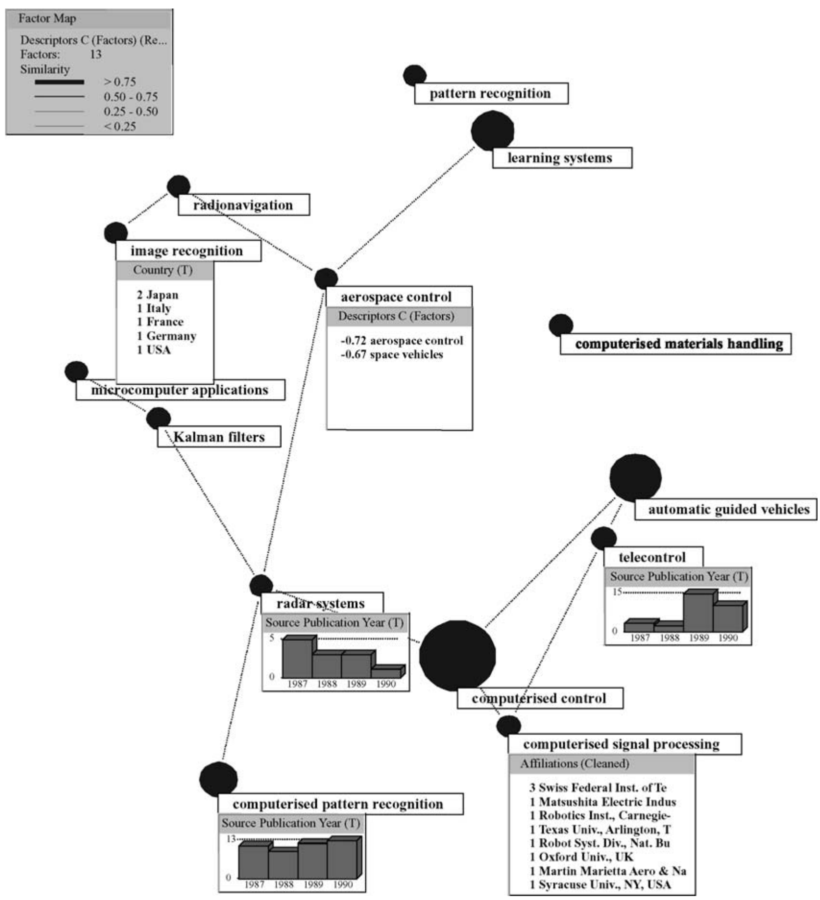

Fig. 1. Autonomous navigation (*INSPEC* 1987–1990) descriptors' standard factor map.

from). The size of a node represents the number of records containing one or more of the high-loading keywords for that factor—e.g., fewer records relate to the "aerospace control" factor than to the "computerized control" factor (lower center in Fig. 1). Location of the factors in the map is based on multidimensional scaling (MDS); it provides a weak reflection of the extent of relationship among factors. The lines connecting factors reflect a path-erasing algorithm; the presence of a connecting line is a stronger reflection of relationship than is map node placement. So, for instance in Fig. 1, the "radar systems" factor (lower left) is somewhat related to four other factors, whereas the "computerized materials handling" factor (upper right) is less related to other factors. *TECH OASIS* (*VantagePoint*) can zoom in to provide various descriptions of a given factor. For instance, in Fig. 1, pull down lists have been frozen in place to illustrate country of the lead authors of the articles relating to the factor "image recognition," year of publication for the articles pertaining to "radar systems," and affiliation of the lead author for articles linked to "computerized signal processing."

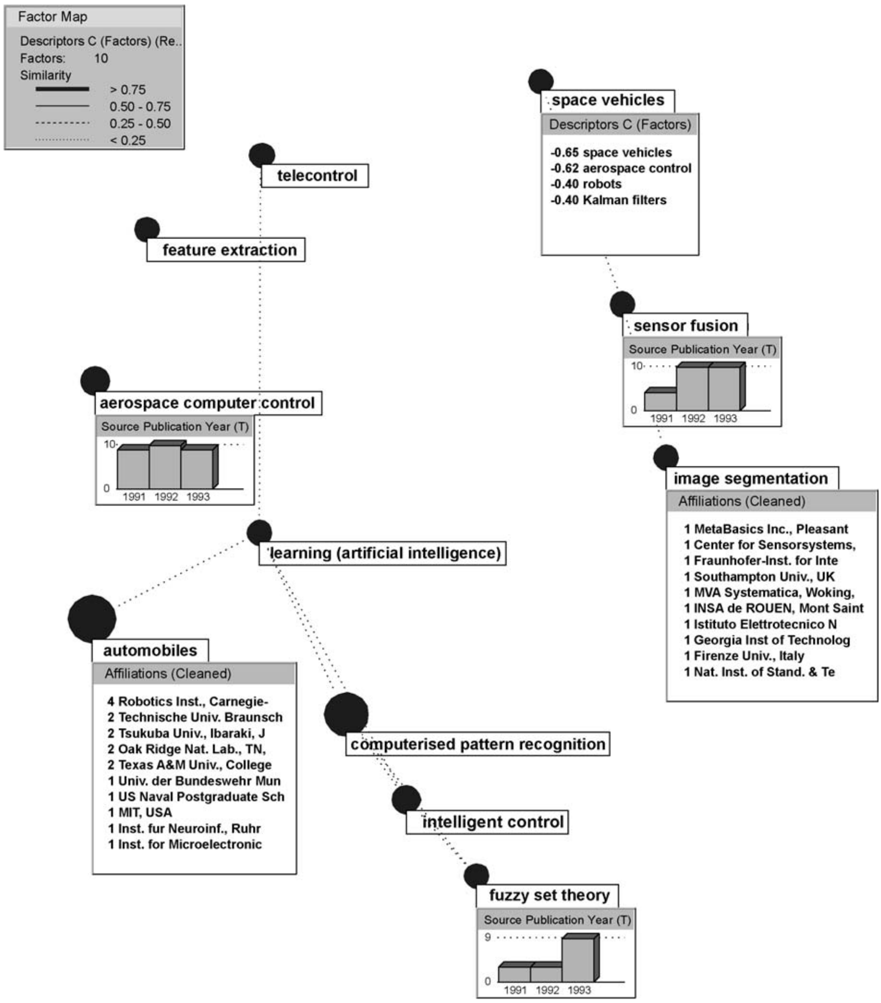

Fig. 2. Autonomous navigation (*INSPEC* 1991–1993) descriptors' standard factor map.

Factor analysis research seeks to create "highly internally homogenous groups, the members of which are similar to one another, and highly externally heterogeneous groups, members of which are dissimilar to those of other groups" [15]. Steinbach et al. [16] discuss and apply measures of cluster quality, both internal and external measures of "goodness." Internal measures, such as cohesion, assess sets of clusters without knowledge of external cluster relationships. External quality measures, such as entropy and F measure, compare factors to known classes, which extends to mean other factors.

We apply an automated process that evaluates factors for each period's publications based on the combined cohesiveness, entropy, and F measure of the derived factor groups [17]. This standard (representing the optimized cluster quality grouping) approach strives to minimize the entropy and F measure, and maximize cohesiveness, for each period's keywords factors.

The factor groups' weighted average entropy, F measure, and cohesion measures for each period will later be plotted across time periods (Fig. 3). Empirically derived relationships among these measures may prove valuable in tracking domain knowledge expansion, as well as technological innovation and diffusion.

Ideally, an assessment of the implications of change in the values of the entropy, F measure, and cohesion would be focused on common factors that reoccur in multiple time periods (e.g., "telecontrol" and "computerized pattern recognition" appear in both Figs. 1 and 2). Evaluation of individual factors' quality measures across periods might reveal specific domain knowledge expansion. Such a situation (i.e., obtaining common factors across periods) seldom occurs naturally. Use of thesauri to seed, or at least encourage, common factors across sequential time periods will be addressed in the discussion on future research. Chen et al. [18] recognize the cluster grouping change issue and state "Base maps across different time intervals tend to have different topology. . . . Such a design tends to give the viewer a relatively high cognitive load because one has to compare different shaped base maps across different time intervals." Only two factors (those just noted) have common titles in both factor maps (Figs. 1 and 2). This does not imply that the other research categories of the 1987–1990 period all ended. However, term usage (i.e., "descriptor" field term/phase record frequencies) changed sufficiently to alter the term/phrase factor analysis factors and names in the subsequent periods. (Note: The term or phrase with the highest loading coefficient is used to name each factor.)

This research focuses on the standard composite (i.e., macrolevel) cluster quality measures derived for the set of factors in each period, as opposed to the individual factors' quality measures. We focus on macrolevel cluster quality measures in hopes that the findings might both validate the standard process, itself, and reveal an approach for individual factor groups' recombination and analysis over sequential time periods.

The following section defines the factor quality measures and addresses their change implications for each period's R&D abstracts. The quality measures will also be assessed as to their empirical relationships to "domain knowledge" and "technology diffusion."

4. Cluster quality measures and change implications

4.1. Cohesion

The internal quality measure applied toward developing a standard factor analysis approach is cohesion. Cohesiveness emanates from the vector space model of document information cluster analysis. In the vector space model, a term frequency vector represents each document. The terms chosen to represent the documents from the "autonomous navigation" abstract set are the leading (most frequent) descriptors. Each descriptor occurs only once in each document. All document vectors, therefore, consist of a sequence of 1's and 0's, as representation of inclusion or exclusion of the descriptors used. Each document vector is normalized to be of unit length. The average pairwise similarity between each factor's documents constitutes the cohesion measure. The pairwise similarity is computed by the vector cosine measure, which for unity vectors equals the vectors' dot product. The standard factor analysis process strives to maximize the factors' cohesion.

What might changes over time in the cohesion measure indicate? A composite cohesion measure (i.e., averaged over each of the factors in each time period) decrease over time might reflect domain knowledge expansion, or a general broadening of the field of research in each subarea (factor). As a technology matures, one might expect domain knowledge to expand and clustered research to become more dissimilar. The opposite change, an increase of the composite cohesion measure, might indicate focusing of domain knowledge, such as on an important new discovery. Knowledge growth could occur in either case. The composite cohesion, therefore, would not seem to be a straightforward indicator for knowledge growth.

4.2. Entropy

Entropy provides an external measure of cluster quality for nonnested clusters or clusters at one level of a hierarchical grouping. The probabilities, P_{ij}, are computed for each cluster grouping. These represent the probability that a member of cluster j belongs to group i, which is defined as the noncommon derived factors. These probabilities can be obtained by analyzing the *TECH OASIS* co-occurrence matrix, which has the derived factors as both the rows and column entries (e.g., Tables 2–4). Cluster group entropy is calculated using the formula:

$$\text{Entropy}_j = -\sum_{i=1}^{m}[P_{ij}\log(P_{ij})]$$

where the sum is taken for all groups, excluding each group where $i=j$. The sum of the weighted entropies for each cluster grouping equals the total entropy:

$$\text{total entropy} = \sum_{j=1}^{m}[(n_j \times \text{Entropy}_j)/n]$$

where n_j equals the number of abstracts in cluster j, m is the number of factors, and n equals the total number of abstracts in the file (e.g., 279 for the 1987–1990 period).

The exclusion of the matrix diagonal entries from the analysis attempts to minimize the comparative entropy penalty that a larger number of factor groups would have vs. a smaller number of factor groups. The applied algorithm attempts to minimize the total factor grouping entropy. However, groupings that generate a large number of factors should not be unduly penalized, since a larger number of small factors may have a higher total cohesion than a smaller number of larger factors. It should be emphasized, the algorithm attempts to maximize total cohesion, while minimizing total entropy, to define the standard factor grouping.

Entropy measures relatedness among factors. What might changes in the composite factors' entropy indicate over sequential periods? A global topic focus, the use of common base technologies, and/or increasing knowledge diffusion might increase the relatedness of common term usage of the constituent factors. As a technology matured, one would expect that base knowledge would be more commonly shared among factors (research clusters), thus increasing the measured entropy. Stated differently, as a technology matures, research papers would become more systems oriented, rather than subtechnology focused, and would be clustered in multiple research factor groups that increasingly overlap. Conversely, if there were a significant discovery or change in research direction in one or more cluster categories, causing divergence in research terms usage, entropy would decrease. More succinctly, convergent research categories across periods would cause the composite entropy to increase. Divergent research categories across periods would lead to lower composite entropy.

4.3. *F* measure

The *F* measure represents the second external cluster quality measure that gets integrated into the standard factor grouping process. The total *F* measure for a factor cluster grouping is defined as

$$F = \sum_{j=1}^{m} [(n_i/n)\max\{F(i,j)\}]$$

where

$$F(i,j) = (2 \times \mathrm{Recall}(i,j) \times \mathrm{Precision}\ (i,j))/(\mathrm{Precision}(i,j) + \mathrm{Recall}(i,j))$$

and

$$\mathrm{Recall}(i,j) = n_{ij}/n_i$$

$$\mathrm{Precision}(i,j) = n_{ij}/n_j$$

n_{ij} equals the number of members of group i in cluster j, n_j is the number of members of cluster j, n_i equals the number of members of group i, and n is the number of documents. As with the entropy calculations, the diagonal values are excluded from the analysis. Again, the standard factor analysis process attempts to minimize the total *F* measure and the total entropy, while maximizing the total cohesion of the derived factor groupings.

The *F* measure represents the maximum similarity—relatedness—between each factor and any of the other factors derived for a period. The increase of *F* measure from one period to the next depicts a significant rise in similarity of one group to at least one, and possibly many other, factors. Both entropy and *F* measure depict external factor groups relatedness. However, the *F* measure provides a composite indicator of the factor groups' maximum relatedness; whereas, total entropy reflects a weighted average of the total intergroup relatedness. If the *F* measure increases and the rate-of-change exceeds that of the entropy, one might suspect that a base factor group(s), mutually common to all factor groups, might have emerged. However, if the total entropy rate-of-change (i.e., the weighted average of the total intergroup relatedness) exceeds that of the *F* measure, clusters of factors may be forming due to general knowledge diffusion. The relevance of *F* measure changes, then, might best be determined by comparative analysis to changes in entropy.

5. Expected patterns and hypotheses

The standard factor analysis strives to minify the effects of factor size (i.e., the number of records in each group) and the number of factors derived as the standard for each period. The composite quality measures for each period equal the weighted average of the individual factors' quality measures divided by the number of factors. The quality criterion are, therefore, normalized to a per factor measure. A summary of the "change implication" discussion for the normalized quality measures includes these hypotheses:

1. Cohesion reduction over periods, conceptually, represents domain knowledge expansion.
2. Cohesion increases over periods, theoretically, depicts domain knowledge focus.
3. Significant increases in entropy per group (i.e., convergent or common research categories) might result from a global subject focus, application of common base technologies, and/or knowledge diffusion.
4. Periods of lower entropy per group might result from or depict a "domain" (i.e., group specific) new discovery or "hot topic" focus (i.e., divergent research between research categories).
5. If the *F* measure rate of increase exceeds that of the entropy, there may be a common related category (i.e., global focus) of the existing factors.
6. If the entropy rate of increase exceeds that of the *F* measure, clusters of factors may be forming due to general knowledge diffusion.

If the factor quality measures can be "properly" normalized/weighted in relation to one another:

1. Periods of higher cohesion vs. entropy might reflect periods of focused parallel research and development.
2. Periods of higher entropy vs. cohesion might reflect periods of knowledge diffusion (i.e., base knowledge multiple factor group applications).

6. Analysis of autonomous navigation R&D evolution

6.1. Basic research: analysis of INSPEC records

As laid out in the Present Research Case section, we analyze the descriptors (keywords) from the *INSPEC* R&D abstracts for successive time periods. The selection of how many descriptors to be analyzed was based on a Zipf distribution analysis [17]. That is, descriptors occurring in the most records were included, with a minimum of 60 descriptors for each time period.

The factors derived for these descriptors were automatically tabularized and depicted in a standard factor map display representation (e.g., Figs. 1 and 2). As discussed, the developed standard (i.e., cluster group optimization) approach applied a composite metric, which minimized the entropy and *F* measures and maximized cohesiveness for each period's

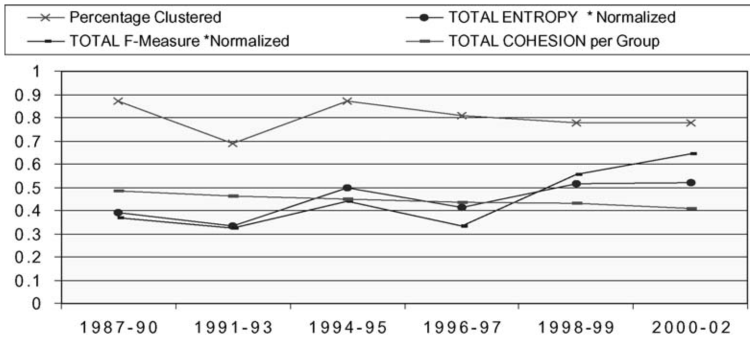

Fig. 3. Autonomous navigation (*INSPEC*) factor groups' composite quality measures evolution.

R&D abstracts. The standard composite entropy, *F* measure, and cohesion measures appear in Table 1, and are plotted across time periods (Fig. 3) to assess their changes in respect to the empirical relationships postulated for domain knowledge expansion or technology diffusion.

In Table 1, the second column shows the number of factors that generated the composite measure. The third column provides the percentage of each period's R&D abstracts that have been included in one or more of the derived standard factors. For example, the 13 factors, which defined 14 cluster groups displayed in Fig. 1, include and/or depict 87% of the 279 abstracts of the 1987–1990 published papers. Similarly, Fig. 2 displays the 10 factors' 11 cluster groups that depict 69% of the 248 abstracts of the 1991–1993 published papers. Columns 4, 6, and 8 in Table 1 list the total entropy per cluster group, the total *F* measure per cluster group, and the total cohesion per cluster group, respectively. The 14 cluster groups of Fig. 1, therefore, have a composite total entropy per group of 0.1077, a composite total *F* measure per group of 0.0272, and a composite total cohesion per group of 0.4849. Similar factor maps to Figs. 1 and 2 were generated for the other periods. The weightings for columns 5 and 7 in Table 1 equal the ratio of the arithmetic mean of column 8 to the arithmetic mean of columns 4 and 6, respectively.

Fig. 3 displays the plots for columns 5, 7, and 8 from Table 1. The total composite cohesion for the period-derived factors declines for the full period analyzed—1987–2001.

Table 1
Autonomous navigation (*INSPEC*) factor groups' composite quality measures

Period	Number of factors	Percentage clustered	Total entropy per group	Total entropy (normalized)	Total *F* measure per group	Total *F* measure (normalized)	Total cohesion per group
1987–1990	13	0.87	0.1077	0.3931	0.0272	0.3691	0.4849
1991–1993	10	0.69	0.0909	0.3316	0.0240	0.3257	0.4621
1994–1995	17	0.87	0.1366	0.4985	0.0324	0.4389	0.4490
1996–1997	14	0.81	0.1130	0.4124	0.0246	0.3339	0.4350
1998–1999	13	0.78	0.1409	0.5140	0.0411	0.5565	0.4296
2000–2002	9	0.78	0.1422	0.5188	0.0476	0.6443	0.4078

This implies that domain knowledge is expanding (i.e., the internal group records are becoming more dissimilar over time). The composite entropy and *F* measure per cluster group declined from the 1987–1990 period to the 1991–1993 period. A significant and comparable rise in both measures occurs for the 1994–1995 period, followed by declines for the 1996–1997 R&D abstracts. The external quality measures, entropy and *F* measure, decline and rise cycle repeats for the 1996–1997 and 1998–1999 periods. During the last two periods, the *F* measure increases significantly, while entropy increases slightly for the 2000–2001 period. Overall, the total entropy per group rises from 0.1077 to 0.1422 during the six periods analyzed. Do the linear regression slopes for the per-group cohesion and entropy calculations from Fig. 3 provide a measure of domain knowledge expansion and technology diffusion for the autonomous navigation basic research?

The *INSPEC* database contains R&D abstracts that generally reflect more basic research than those compiled in *EI Compendex*. From Fig. 3 and the earlier hypotheses, one might surmise that the autonomous navigation basic research for the period between the 1987–1990 and 1991–1993 was internally group focused (i.e., divergent research based on the lower entropy calculations) with each domain's (i.e., factor group) knowledge expanding, as

Table 2
Autonomous navigation standard *INSPEC* factors co-occurrence matrix for 1994–1995

No. of records	Description C (17 factors)								
	No. of records (306)	71	63	46	45	37	36	31	28
	6 (1994-1995)	Digital simulation	Robot vision	Aerospace computing	Image sequences	Road vehicles	Traffic control	Inference mechanisms	Neuro-controllers
71	Digital simulation	71	18	14					9
63	Robot vision	18	63	26	14	9	12		
46	Aerospace computing	14	26	46	10				
45	Image sequences		14	10	45	9	9		
37	Road vehicles		9		9	37	9	7	
36	Traffic control		12		9	9	36	9	
31	Inference mechanisms					7	9	31	
28	Neurocontrollers	9							28
26	Fuzzy logic								
26	Control system synthesis								8
21	Self-organizing feature maps								
21	Inertial navigation								
18	Image texture				14				
17	Space research	9		17					
17	Transportation						6	10	
16	Marine systems								
16	Planning (artificial intelligence)								
14	Cameras		7	14					
10	Parallel processing								

depicted by the lower cohesion per group calculation. Between the 1991–1993 and the 1994–1995 periods, the *INSPEC* basic research factors' entropy increased, suggesting a global subject focus and/or knowledge diffusion. To determine which, if either, of these events (i.e., global subject focus or knowledge diffusion) have occurred, we shall assess the factors' common records.

Table 2 presents the co-occurrence matrix of the records contained in the 1994–1995 factors. Note the occurrence of clusters of factor groups with common records (i.e., the shaded areas). The duplicate records within multiple groups occur because of common usage of the factor–defining descriptors across the represented research clusters. This shared research documentation across categories appears to represent knowledge diffusion or convergent research, more so than the application of common base or focus technology or application. The clusters of factors (i.e., convergent research) may be analogous to the pieces of a puzzle coming together; depicting the formation of subdisciplines within the technology. If so, "road vehicles," "traffic control," and "inference mechanisms" would depict such a subdiscipline. Two factors, (truncated from Table 2 and not shown), "space research" and "cameras," are subsets of the group "aerospace computing." The existence of these subsets causes the derived entropy and F measure calculations to increase more than might be expected, thus skewing the entropy and F measure points in Fig. 3 higher than, perhaps, they should be.

Moving forward two periods, the entropy and F measure increases during the 1998–1999 period appear to be due to the factor group "computerized navigation." Note that the F measure rate of change exceeds that for the entropy calculation. Table 3 shows the group "computerized navigation" to have common records with most of the other factor groups. Does the research documented under the "computerized navigation" group represent a

Table 3
Autonomous navigation standard *INSPEC* factors co-occurrence matrix for 1998–1999

No. of records	Descriptors C (13 Factors)					
	No. of records (284)	91	39	38	35	33
	6 (1998-1999)	computerised navigation	inertial navigation	road traffic	CCD image sensors	remotely operated vehicles
91	computerised navigation	91	28	8	12	
39	inertial navigation	28	39			
38	road traffic	8		38	6	
35	CCD image sensors	12		6	35	
33	remotely operated vehicles					33
32	inference mechanisms	17		7	6	
29	image sequences	9	14	7		
28	robot kinematics	9				
27	neurocontrollers	20				
25	image matching	6				13
23	space vehicles	8				
19	object detection	6				
13	neural nets	9				

Table 4
Autonomous navigation standard *INSPEC* factors co-occurrence matrix for 2000–2001

No. of Records	Descriptors C (9 FACTORS)									
	No. of Records (223)	74	46	35	34	29	27	25	19	14
	4 (2000–2001)	Image sequences	Robot vision	Kalman filters	Auto-mobiles	Image sensors	Optimal control	Aircraft control	Control system synthesis	Space vehicles
74	Image sequences	74	18	10	25	12	7	6	4	6
46	Robot vision	18	46	4	8	8		5		
35	Kalman filters	10	4	35		6	5	5		
34	Automobiles	25	8		34	7				
29	Image sensors	12	8	6	7	29				
27	Optimal control	7		5			27	6	4	
25	Aircraft control	6	5	5			6	25	4	
19	Control system synthesis	4					4	4	19	4
14	Space vehicles	6						4		14

base or focus for the other research factors? This factor's descriptors—computerized navigation, sensor fusion, fuzzy control, fuzzy logic, image sensors, distance measurement, uncertainty handling, and software agents—provide a hint as to the overall subject matter of the documented research. The clusters of factors (i.e., shaded areas) do not appear as prevalent for the 1998–1999 factor groups (Table 3) as for the 1994–1995 factor groups (Table 2).

Total entropy per group rises only slightly from 0.1408 to 0.1422 between the last two time periods (Table 1). For the 2000–2001 period, Table 4 shows both a high entropy group, "image sequences," and clusters of factors. At first, one might question the dissimilar base or focus groups shown in Tables 3 and 4; however these groups' difference may be largely in name. In fact, the most frequent group-defining term of the "image sequences" group (Table 4) is "computerized navigation."[2] Comparing the base or focus factors' group defining terms, the research emphasis also appears to have evolved from component level research to a system level emphasis. For example, the 1998–1999 group-defining term, "image sensors," is a component of the 2000–2001 group-defining term, "computer vision," a system level descriptor. A similar relationship exists for the 1998–1999 descriptor, "distance measurement," and the 2000–2001 descriptor, "motion estimation." Note in Table 4 that there appears to be a secondary focus group, "aircraft control," an even more specific application than "computerized navigation."

[2] The complete list of group defining terms for "image sequences" includes computerized navigation, computer vision, real-time systems, vehicles, image sequences, and motion estimation.

Table 5
Autonomous navigation (*EI Compendex*) factor groups' composite quality measures

Period	Number of factors	Percentage clustered	Total entropy per group	Total entropy (normalized)	Total *F* measure per group	Total *F* measure (normalized)	Total cohesion per group
1987–1990	11	0.87	0.1292	0.3260	0.0380	0.4124	0.4758
1991–1993	10	0.89	0.1639	0.4137	0.0531	0.5757	0.4203
1994–1995	15	0.89	0.1806	0.4557	0.0314	0.3405	0.3864
1996–1997	21	0.93	0.1763	0.4450	0.0311	0.3371	0.3975
1998–1999	12	0.82	0.1187	0.2996	0.0250	0.2709	0.4114
2000–2002	13	0.89	0.2176	0.5491	0.0510	0.5525	0.3977

6.2. Applied research: analysis of EI compendex records

To determine whether the "innovation indicator" implications of the "factor quality measures' changes" can be more globally applied, the 1091 autonomous navigation abstracts from *EI Compendex* were separately analyzed. These reflect more applied research than the *INSPEC* records, so this assessment addresses a somewhat more mature stage in the autonomous navigation research and development. As with the *INSPEC* abstracts, the *EI Compendex* abstracts were subdivided into roughly equal-record periods and subjected to the same type of analysis. The standard composite measures for entropy, *F* measure, and cohesion were tallied (Table 5) and plotted over the six time periods (Fig. 4).

Unlike the more basic research (*INSPEC* record analyses), the factors' entropy and *F* measure increase significantly over the periods from 1987–1990 to 1991–1993. Initially, the rate of increase for the composite *F* measure exceeds that of the composite entropy. This is similar to the relationship observed for the *INSPEC* 1998–1999 and 2000–2001 periods. As observed in Tables 3 and 4, Table 6 also shows a base or focus group, "velocity control," and the clusters of factor groups (i.e., shaded areas). The composite factors' cohesion decreases over the first three periods, depicting domain knowledge

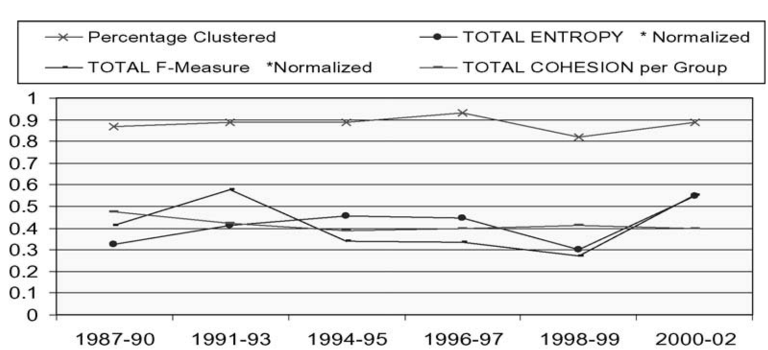

Fig. 4. Autonomous navigation (*EI Compendex*) factor groups' composite quality measures evolution.

Table 6
Autonomous navigation (*EI Compendex*) standard factors co-occurrence matrix for 1991–1993

No. of Records	Descriptors C (10 FACTORS)												
	No. of Records (164)	39	37	36	32	29	28	19	19	18	14	13	8
	4 (1991-1993)	Pattern recognition	Velocity control	Mobile robots	Lasers	Algorithms	Robots, Industrial	Learning systems	Satellites	Performance	Electronic guidance systems	Underwater equipment	Range finders
39	Pattern recognition	39	5	6	5	12	10		4				
37	Velocity control	5	37	17		7		5	7	8			
36	Mobile robots	6	17	36		10				5		10	
32	Lasers	5			32	5	11	4					
29	Algorithms	12	7	10	5	29	7	4					
28	Robots, Industrial	10			11	7	28			5			
19	Learning systems		5		4	4		19		4	4		
19	Satellites	4	7						19				
18	Performance		8	5			5	4		18			
14	Electronic guidance systems							4			14		
13	Underwater equipment		10									13	
8	Range finders												8

Table 7
Autonomous navigation (*EI Compendex*) standard factors co-occurrence matrix for 1998–1999

No. of records	Descriptors C (12 FACTORS)						
	No. of records (182)	28	27	26	25	24	21
	4 (1998-1999)	Intelligent control	Intelligent vehicle highway systems	Fuzzy control	Real–time systems	Image analysis	Acoustic devices
28	Intelligent control	28		7	4		
27	Intelligent vehicle highway systems		27			6	
26	Fuzzy control	7		26			
25	Real-time systems	4			25		
24	Image analysis		6			24	
21	Acoustic devices						21
20	Mathematical models						
19	Closed-loop control systems	6	5				4
18	Charge-coupled devices				4		
18	Vehicles			5			
15	Video cameras						
13	Sensor data fusion						
11	Constraint theory						

Table 8
Autonomous navigation (*EI Compendex*) standard factors co-occurrence matrix for 2000–2001

No. of records	Descriptors C (13 FACTORS)						
	No. of records (180)	45	43	39	36	34	34
	4 (2000–2001)	Intelligent robots	Intelligent vehicle highway systems	Satellites	Vehicle wheels	Bombs (ordnance)	Dynamic programming
45	Intelligent robots	45	10	8	7	9	11
43	Intelligent vehicle highway systems	10	43	10	8	8	9
39	Satellites	8	10	39	9	8	13
36	Vehicle wheels	7	8	9	36	16	6
34	Bombs (ordnance)	9	8	8	16	34	6
34	Dynamic programming	11	9	13	6	6	34
26	Feature extraction		5		4	8	4
23	Computational geometry	4	8		7	5	4
21	Closed–loop control systems	7	4	4			7
20	Fuzzy control			6		4	
15	Oceanography		4				
8	Motion estimation						
8	Learning algorithms		7				

expansion. Stated in different terms, the similarity among research factors increases, as the similarity of the research within factors decreases. The global emphasis appears to be on mobile or industrial robots.

Between the 1994–1995 and 1998–1999 periods, the factors' composite entropy and *F* measure decline. The rate of change of the composite entropy exceeds that of the composite *F* measure. Table 7 shows that neither a base factor nor clusters of factor groups exist for the low entropy period, 1998–1999. Both external cluster group relatedness measures then rise significantly in the 2000–2002 period. The steep increase of the entropy and *F* measure in the 2000–2001 period results from the reemergence of a base factor group, "intelligent vehicle highway systems" (IVHS), and clusters of factor groups, as shown in Table 8. Note the focus shift from "mobile or industrial robots" to the IVHS. Does the decline and rise cycle of entropy and *F* measure signal a shift in research focus? The composite cohesion declines in all periods except for 1996–1997 and 1998–1999. As with the full cycle changes of the *INSPEC* cluster group quality measures, the *EI Compendex* factor grouping cohesion measure decreased and entropy increased. However, more period-to-period variations occur in the *EI Compendex* abstracts' analysis, as might be indicative of a research focus change.

7. Additional observation

What else has been revealed? The entropy increase observed for the *INSPEC* and *EI Compendex* factors would seem to relate to a cluster group "degree of focus." Note in Table 9, the degree of focus (i.e., how specific the factors' subjects appear) does seem to be directly related to the entropy. Table 9 lists the high entropy groups derived for the *EI Compendex* autonomous navigation abstracts for each period. The high entropy groups would have the greatest effect on the composite entropy calculation. Observe how generic the first set of factor group names appear (e.g., "systems science and cybernetics," "computer software," etc.). Entropy per group was low for this period (1987–1990). The next period's group names, those for 1991–1993, appear more specific (e.g., "pattern recognition," velocity control," "mobile robots," etc.). Entropy increased during this period (1991–1993).

There is also a direct relationship between entropy per group and the percentage of the abstracts included in the factors (i.e., standard factor map). Columns 3 (percentage clustered) and 4 (total entropy per group) of Tables 1 and 5 depict this relationship. By comparing the percentage clustered between Table 1, which displays analysis results for more basic research, and Table 5, providing the analyses of the more applied research, one sees that a greater percentage of the applied research abstracts get clustered than do for the basic research. The *INSPEC* percentage clustered ranges from a low of 69% clustered to a high of 87% clustered, with an average of 80%. The *EI Compendex* "percentage clustered" ranges from a low of 82% clustered to a high of 93% clustered, with an average of 88% (Tables 1 and 5). Most will accept the premise that basic research is less focused than applied research. These observations and relationships would

Table 9
Autonomous navigation (*EI Compendex*) high entropy factors

Group names	Group entropy	Group cohesion	Group names	Group entropy	Group cohesion
1987–1990: Systems science and cybernetics (48)	0.4604	0.4460	1996–1997: Bandwidth (42)	0.4725	0.3417
1987–1990: Computer software (33)	0.2447	0.3980	1996–1997: Kalman filters (36)	0.4503	0.3660
1987–1990: Image analysis (35)	0.2391	0.5338	1996–1997: Intelligent vehicle highway systems (33)	0.3664	0.4402
1987–1990: Navigation aids application (27)	0.1955	0.5047	1996–1997: Membership functions (28)	0.2539	0.4275
1991–1993: Pattern recognition (39)	0.3089	0.4444	1996–1997: Feature extraction (27)	0.2519	0.4467
1991–1993: Velocity control (37)	0.3083	0.3389	1996–1997: Obstacle detectors (26)	0.2470	0.3606
1991–1993: Mobile robots (36)	0.2785	0.4673	1996–1997: Genetic algorithms (25)	0.2377	0.3629
1991–1993: Algorithms (29)	0.2404	0.4788	1996–1997: Charge-coupled devices (23)	0.2299	0.4031
1991–1993: Lasers (32)	0.2021	0.4330	1996–1997: Optical sensors (26)	0.2033	0.3602
1991–1993: Robots, industrial (28)	0.1649	0.5195	1998–1999: Intelligent control (28)	0.2023	0.4503
1994–1995: Control system analysis (42)	0.3610	0.4261	1998–1999: Intelligent vehicle highway systems (27)	0.1975	0.4100
1994–1995: Automobile electronic equipment (36)	0.3055	0.4079	1998–1999: Image analysis (24)	0.1639	0.3939
1994–1995: Numerical methods (36)	0.2767	0.3791	1998–1999: Real-time systems (25)	0.1630	0.4339
1994–1995: Robotic arms (32)	0.2683	0.3697	1998–1999: Fuzzy control (26)	0.1530	0.4093
1994–1995: Recursive functions (29)	0.2468	0.4066	2000–2001: Intelligent robots (45)	0.4184	0.4066
1994–1995: Real-time systems (30)	0.2417	0.4144	2000–2001: Intelligent vehicle highway systems (43)	0.3919	0.4208
1994–1995: Cameras (29)	0.2238	0.3931	2000–2001: Vehicle wheels (36)	0.3227	0.3451
1994–1995: Fuzzy control (27)	0.1973	0.4026	2000–2001: Dynamic programming (34)	0.3217	0.3747
			2000–2001: Bombs (ordnance) (34)	0.3185	0.3694
			2000–2001: Satellites (39)	0.3120	0.3784

then support the premise that the composite entropy per group reflects a degree of focus of the clustered research.

In technology management, R&D focus was found to be a factor for successful innovation (i.e., at the organizational level of R&D management) [19]. This composite entropy calculation might then provide a measure to extend the "degree of focus"

Table 10
Autonomous navigation (*INSPEC*) 1987–2002 record-periods factors recombination groups

Group 6 (Recombo-4: image sequences)	Group 5 (1998–1999: computerized navigation)	Group 4 (Recombo-3: Kalman filters)
1991–1993: image segmentation	1991–1993: sensor fusion	1994–1995: marine systems
1994–1995: digital simulation	1991–1993: fuzzy set theory	1987–1990: microcomputer applications
1998–1999: image matching	Recombo-2: fuzzy logic	1996–1997: helicopters
1994–1995: image texture	2000–2002: image sensors	2000–2002: aircraft control
1996–1997: image colour analysis	1998–1999: neural nets	2000–2002: optimal control
	Recombo-2: neurocontrollers	1996–1997: edge detection
	Recombo-2: control system synthesis	Recombo-2: image recognition
	Recombo-2: inertial navigation	1996–1997: co-operative systems
	Recombo-2: robot kinematics	

Group 7 (Recombo-2: robot vision)	Group 8 (Recombo-2: road traffic)	Group 1 (1987–1990: computerized control)
1991–1993: feature extraction	1994–1995: transportation	1987–1990: radionavigation
1998–1999: CCD image sensors	1994–1995: road vehicles	1987–1990: radar systems
1994–1995: aerospace computing	Recombo-2: inference mechanisms	1987–1990: learning systems
1994–1995: cameras	1994–1995: traffic control	1987–1990: computerized materials handling
	1998–1999: object detection	1987–1990: pattern recognition
	1994–1995: self-organising feature maps	1987–1990: automatic guided vehicles
	1994–1995: planning (artificial intelligence)	Recombo-2: telecontrol
		1994–1995: parallel processing

Table 10 (*continued*)

Group 3 (Recombo-3: space vehicles)	Group 2 (Recombo-2: computerized pattern recognition)
Recombo-2: aerospace control	1991–1993: intelligent control
1994–1995: space research	1991–1993: learning (artificial intelligence)
1998–1999: remotely operated vehicles	1991–1993: aerospace computer control
1996–1997: feedback	1987–1990: computerized signal processing
1996–1997: virtual reality	Recombo-3: automobiles

assessment of successful innovation to the next source level (e.g., industry segment, nation, whatever).

8. Future research

The cluster quality measures for the individual factors (e.g., shown in Table 9) serve to derive the composite quality measures for each factor grouping, as well as factor grouping optimization for any given period. We hope to extend the analysis to evaluation of the individual cluster group level. To do so, common or linked factors across periods are necessary. Such a situation (i.e., obtaining common factors across periods) could be seeded, or at least encouraged, by selecting the "subject specific keywords" to be analyzed. If a thesaurus of terms and phrases specific to the subject matter (i.e., in this evaluation, autonomous navigation) was available, it could be used to tag the "subject" relevant terms to include in the cluster analyses. These "relevant" terms in the combined file could be used to tag the "most relevant" terms in the files containing the abstracts from each period analyzed. However, even with term seeding, common factors across time periods do not always occur due to the changing research emphases, as reflected in the descriptors/keywords record frequencies. The greater value that may be derived from usage of subject matter specific thesauri may be in the generation of more relevant, domain-specific, factors.

Common factors across time periods do occur naturally without "subject matter term seeding," but less frequently than might be desired for a subtopic specific analysis (e.g., on "telecontrol" or "computerized pattern recognition" that appear in both Figs. 1 and 2). An automated analysis process that uses a relatedness assessment of factor "group defining terms," as well as cross-group record commonality, has been developed to link factors across the analyzed periods. Table 10 shows the recombined factors for each period's derived factor groups for the 1629 *INSPEC* "autonomous navigation" research literature abstracts. In Table 10, group names appended with "Recombo-#*X*" represent factors with identical names that occurred in #*X* of the record periods. For example, "Recombo-2: robot vision" indicates that the factor named "robot vision" was derived in two of the six record periods. Observe in Group 3 of Table 10 that the previously discussed groups, "1987–1990: aerospace control" and "1991–1993: space vehicles," have been recombined into a group titled "space vehicles." Observe in Fig. 2 that the group "1991–1993: space vehicles" has a "high-factor-loading descriptor" (i.e., group defining term) of "Kalman filters." Had the group "space vehicles" not occurred in two other periods' factor groupings, one might argue that the group "1991–1993: space vehicles" could be combined with the "Kalman filters" recombination group, Group 4 in Table 10. Obviously, research is not uniquely restricted to single categories. However, to understand the primary evolution of the subcategories of a field of research, the automated recombination algorithm attempts to uniquely recombine period-derived factors. Due to the discussed ambiguities, the recombination process is still under development and in need of subject-matter expert critique of combined categories.

Another approach to time-slice R&D analysis would be to use document-oriented mapping. We are exploring integration of another software package (e.g., *VX Insight*). One form of mapping develops clusters of similar documents based on co-occurrence of terms, as described herein, but to show document clusters rather than term clusters (factors). Such maps can be generated for an entire time period. Then, documents for particular time periods can be plotted as colored overlays on the full map. This is one way to overcome the difficulty of dealing with dissimilar factors over time periods.

9. Conclusions

Research continues on developing a time-slice recombination process to link "similar," but differently named factors across periods. However, the recombination process is still experimental. Therefore, this paper focuses on the use and change implications of the composite factor set quality indices. To do so, we introduce and apply the *TECH OASIS* software system and the PCA-based factor map analysis. We present the factor grouping quality measures: cohesion, entropy, and F measure.

Using the standard factor map analysis process (i.e., that which selects the optimal number of factor groups based on a metric that strives to minimize factors' composite entropy and F measure, while maximizing cohesion), the most relevant terms (descriptors) for each time period are analyzed and clustered. We generate factor maps, such as shown in Figs. 1 and 2. We then plot and assess changes in the factors' composite quality measures against empirical hypotheses relating to the maturation of a technology, specifically domain knowledge expansion and technology diffusion.

Within this limited case study of a particular technology, autonomous navigation, we observe consistent patterns of factor set quality measure changes over the periods analyzed. For both the basic and applied research, as represented by 1629 *INSPEC* and 1091 *EI Compendex* R&D abstracts, respectively, factor sets' cohesion declined and total entropy increased over time. Lower factor cohesion results from each factor group's R&D abstracts becoming more dissimilar (i.e., domain knowledge expansion). Entropy increases as each set of factors has greater commonality of constituent abstracts. Confirming whether the linear regression slopes of the cohesion and entropy measures, as shown in Figs. 3 and 4, serve to measure domain knowledge expansion and technology diffusion requires further and expanded analyses of other technologies.

We note that the relatedness of factor groups (i.e., entropy) can increase for several reasons—a global topic focus, the use of common base technologies, and/or increasing knowledge diffusion with the formation of subdisciplines within the field. Some differentiation of the causes of periodic entropy rises are noted by comparing with the *F* measure. Entropy increasing at a greater rate than the *F* measure appears to depict the formation of clusters of factors (i.e., the formation of subdisciplines within the technology analyzed). *F* measure rates of increase greater than that for entropy appear to signify the formation of a common factor that can represent either a base technology or focal application.

We note the cycling of entropy rise and decline, within autonomous navigation applied research, and observe that the research focus appears to shift from "industrial or mobile robots" to the "intelligent vehicle highway system" as entropy again increases. We propose that entropy can measure the "level-of-focus" of the information analyzed, based on derived factor name specificity and percent of abstract clustered relationships. Whether the entropy measure can assess macro "source" level (e.g., industry sectors, nations) R&D activity patterns to project innovation success requires further research.

We have posed a number of questions regarding the potential of three factor quality measures applied to compilations of R&D abstracts to help assess technology maturation. To answer these questions, we have begun analyses of "automotive lightweight materials" in the 1990s and of "smart materials." However, the most significant outcome of this case study is that the observed logical patterns of factor quality measure trends seem to validate the standard (i.e., cluster quality measures optimization) factor analysis process. A standard factor map process permits such comparative periodic R&D assessments and expands the innovation analysis capabilities of the *TECH OASIS* software system.

References

[1] A.F.J. van Raan, Advanced bibliometric methods to assess research performance and scientific development: basic principles and recent practical applications, Res. Eval. 3 (3) (1992) 151–166 (see also http://sahara.fsw.leidenuniv.nl/cwts/nnmap0.html).

[2] A.L. Porter, A.T. Roper, T.W. Mason, F.A. Rossini, J. Banks, Forecasting and Management of Technology, Wiley, New York, 1991.

[3] A.L. Porter, X.-Y. Jin, J.E. Gilmour, S.W. Cunningham, H. Xu, C. Stanard, L. Wang, Technology opportunities analysis: integrating technology monitoring, forecasting & assessment with strategic planning, Soc. Res. Adm. J. 21 (2) (1994) 21–31.

[4] A.L. Porter, M.J. Detampel, Technology opportunities analysis, Technol. Forecast. Soc. Change 49 (2) (1995) 237–255.

[5] R.J. Watts, A.L. Porter, S.W. Cunningham, D. Zhu, Vantage point intelligence mining: analysis of natural language processing and computational linguistics, in: J. Komorowski, J. Zytkow (Eds.), Principles of Data Mining and Knowledge Discovery, First European Symposium, PKDD'97, Trondheim, Norway, Springer-Verlag, Tiergartenstr 17, D-69121 Heidelberg, Germany, 1997, pp. 323–335.

[6] A.L. Porter, A. Kongthon, J.-C. Lu, Research profiling: improving the literature review, Scientometrics 53 (2002) 351–370.

[7] R.J. Watts, A.L. Porter, C. Courseault, Functional analysis: deriving systems knowledge from bibliographic information resources, Inf. Knowl. Syst. Manage. 1 (1) (1999) 45–61.
[8] R.J. Watts, A.L. Porter, N.C. Newman, Innovation forecasting using bibliometrics, Compet. Intell. Rev. 9 (4) (1998) 1–9.
[9] R.J. Watts, A.L. Porter, Innovation forecasting, Technol. Forecast. Soc. Change 56 (1997) 25–47.
[10] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, D. Harshman, Indexing by latent semantic analysis, J. Am. Soc. Inf. Sci. 41 (1990) 391–407.
[11] J.P. Carlisle, S.W. Cunningham, A. Nayak, A.L. Porter, Related problems of knowledge discovery, Hawaii International Conference on System Sciences (HICSS) Proceedings on CD—Modeling Technologies and Intelligent Systems Track; Data Mining and Knowledge Discovery Mini-Track, January, 1999.
[12] D. Zhu, A.L. Porter, S. Cunningham, J. Carlisle, A. Nayak, A process for mining science & technology documents databases, illustrated for the case of knowledge discovery and data mining, Cienc. Inf. 28 (1) (1999) 1–8.
[13] D. Zhu, A.L. Porter, S.W. Cunningham, J. Carlisle, A. Nayak, A process for mining science & technology documents databases, illustrated for the case of "knowledge discovery and data mining," Internal TOA Paper #94 (available on request).
[14] P. Losiewicz, D.W. Oard, R.N. Kostoff, Textual data mining to support science and technology management, J. Intell. Inf. Syst. 15 (2) (2000) 99–119.
[15] K. Borner, C. Chen, K.W. Boyack, Visualizing knowledge domains, ARIST 37. In Press.
[16] M. Steinbach, G. Karypis, V. Kumar, A comparison of document clustering techniques, Technical Report #00-034, University of Minnesota, 2000 (http://www.cs.umn.edu/tech_reports/).
[17] R.J. Watts, A.L. Porter, D. Zhu, Factor analysis standardization: demonstrated on natural language knowledge discovery, J. Am. Soc. Inf. Sci. Technol. (submitted for publication).
[18] C. Chen, T. Cribbin, R. Macredie, S. Morar, Visualizing and tracking the growth of competing paradigms: two case studies, J. Am. Soc. Inf. Sci. Technol. 53 (8) (2002) 678–689.
[19] W.E. Souder, Managing New Product Innovations, Lexington Books, New York, 1987, pp. 199–216.

Robert Watts currently works for the Tank-Automotive and Armaments Command (TACOM), Advanced Vehicle Technologies, as team leader for technology assessments. He is both managing and participating in the development of the competitive intelligence software system, Technology Opportunities Analysis of Scientific Information System (*TECH OASIS*).

Alan Porter is professor emeritus of industrial and systems engineering and public policy at the Georgia Institute of Technology.